Effect Decomposition in the Presence of an Exposure-Induced Mediator-Outcome Confounder

Tyler J. VanderWeele,[a] *Stijn Vansteelandt,*[b] *and James M. Robins*[a]

Abstract: Methods from causal mediation analysis have generalized the traditional approach to direct and indirect effects in the epidemiologic and social science literature by allowing for interaction and nonlinearities. However, the methods from the causal inference literature have themselves been subject to a major limitation, in that the so-called natural direct and indirect effects that are used are not identified from data whenever there is a mediator-outcome confounder that is also affected by the exposure. In this article, we describe three alternative approaches to effect decomposition that give quantities that can be interpreted as direct and indirect effects and that can be identified from data even in the presence of an exposure-induced mediator-outcome confounder. We describe a simple weighting-based estimation method for each of these three approaches, illustrated with data from perinatal epidemiology. The methods described here can shed insight into pathways and questions of mediation even when an exposure-induced mediator-outcome confounder is present.

(*Epidemiology* 2014;25: 300–306)

One of the main advantages of the counterfactual approach to mediation analysis is allowing for effect decomposition of a total effect into a natural direct and indirect effect even in models with interactions and nonlinearities.[1,2] This is accomplished using new definitions of direct and indirect effects defined in terms of counterfactuals,[1,2] often referred to as natural direct and indirect effects. Robins and Greenland[1] noted that natural direct and indirect effects could not be identified even when both observational and experimental data were available. However, Pearl[2] showed that under additional causal assumptions, encoded in a causal diagram interpreted as a nonparametric structural equation model with independent errors,[2] identification was possible if the following four assumptions held: (i) the effect of the exposure A on the outcome Y is unconfounded conditional on C; (ii) the effect of the mediator M on the outcome Y is unconfounded conditional on C; (iii) the effect of the exposure A on the mediator M is unconfounded conditional on C; and (iv) none of the mediator-outcome confounders are themselves affected by the exposure. Throughout this article, we assume that our causal diagrams represent underlying nonparametric structural equation models.

It has, however, also been shown[3] that if there is an intermediate variable that is affected by the exposure and that in turn confounds the mediator-outcome relationship—thereby violating assumption (iv) above—then, even under an nonparametric structural equation model with independent errors, natural direct and indirect effects are not identified from the data, irrespective of whether data were collected on this intermediate confounding variable. Natural direct and indirect effects are still theoretically appealing, but because they cannot be identified, bounds or sensitivity analysis must be applied. This essentially has restricted the contemporary methods for causal mediation analysis to settings in which the mediator occurs shortly after the exposure to minimize the possibility of such exposure-induced mediator-outcome confounding.[4] This is a severe limitation. It is not one that is possible to address directly.

In this article, we partially circumvent this limitation by providing three approaches to effect decomposition in the presence of such an exposure-induced mediator-outcome confounder. First, we will consider definitions of direct and indirect effects in which the exposure-induced confounder and the original mediator of interest are instead both jointly considered as the mediator. Second, we will consider certain path-specific effects that can be identified from data in settings with an exposure-induced mediator-outcome confounder.[3] Finally, we will consider a randomized interventional analog of the notions of natural direct and indirect effects that can also be identified from data in the presence of an exposure-induced mediator-outcome confounder. These three approaches to effect decomposition—although not estimating the natural direct and indirect effects themselves—may provide some insight into mediation and pathways in this more challenging, though not uncommon, setting. As such, the methods

Submitted 16 May 2013; accepted 20 August 2013.

From the [a]Departments of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, MA; and [b]Department of Applied Mathematics, Computer Science, and Statistics, University of Ghent, Ghent, Belgium.

T.J.V. was supported by National Institutes of Health grant ES017876. S.V. was supported by Flemish Research Council (FWO Grant G.0111.12).

SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com). This content is not peer-reviewed or copy-edited; it is the sole responsibility of the author.

Correspondence: Tyler J. VanderWeele, Departments of Epidemiology and Biostatistics, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115. E-mail: tvanderw@hsph.harvard.edu.

ISSN: 1044-3983/14/2502-0300
DOI: 10.1097/EDE.0000000000000034

presented in this article may help to address one of the major limitations of the causal mediation analysis literature.

DIRECT AND INDIRECT EFFECTS: NOTATION AND DEFINITIONS

Let A denote the exposure of interest, Y the outcome of interest, M the potential mediator of interest, and C a set of baseline covariates not affected by the exposure. We will let Y_a and M_a denote, respectively, the values of the outcome and mediator that would have been observed had the exposure A been set to level a; let Y_{am} denote the value of the outcome that would have been observed had A been set to level a, and M to m. These counterfactual variables, Y_a, M_a, and Y_{am}, all presuppose that at least hypothetical interventions on A and M are conceivable. Some additional technical conditions referred to as consistency and composition are also needed to relate the observed data to counterfactual quantities. The consistency assumption in this context is that when $A = a$, the counterfactual outcomes Y_a and M_a are, respectively, equal to the observed outcomes Y and M and that when $A = a$ and $M = m$, the counterfactual outcome Y_{am} is equal to Y. The composition assumption is that $Y_a = Y_{aM_a}$. Further discussion on these assumptions is given elsewhere.[4,5]

Suppose a and a^* are two values of the exposure we wish to compare, for example, for binary exposure we may have $a = 1$ and $a^* = 0$. The average controlled direct effect comparing exposure level $A = a$ to $A = a^*$ and fixing the mediator to level m is defined by $E[Y_{am} - Y_{a^*m}]$ and captures the effect of exposure A on outcome Y, intervening to fix M to m; it may be different for different levels of m.[1,2] Direct effects are always relative to the mediator M being considered (ie, through pathways other than M). The natural direct effect[1,2] is defined as $E\left[Y_{aM_{a^*}} - Y_{a^*M_{a^*}}\right]$ and differs from controlled direct effects, in that the intermediate M is set to the level M_{a^*}, the level that it would have naturally been under some reference condition for the exposure, $A = a^*$. Similarly, the average natural indirect effect can be defined as $E\left[Y_{aM_a} - Y_{aM_{a^*}}\right]$, which compares the effect of the mediator at levels M_a and M_{a^*} on the outcome when exposure is set to $A = a$. For the natural indirect effect to be nonzero, the exposure would have to change the mediator and that change in the mediator would have to change the outcome; natural indirect effects thus formally capture our notion of mediation. Natural direct and indirect effects are referred to by Robins and Greenland[1] as "pure direct effects" and "total indirect effects," respectively. Natural direct and indirect effects have the property that a total effect, $E[Y_a - Y_{a^*}]$, decomposes into a natural direct and indirect effect:

$$E[Y_a - Y_{a^*}] = E\left[Y_{aM_a} - Y_{a^*M_{a^*}}\right]$$
$$= E\left[Y_{aM_a} - Y_{aM_{a^*}}\right] + E\left[Y_{aM_{a^*}} - Y_{a^*M_{a^*}}\right];$$

the decomposition holds even when there are interactions and nonlinearities.

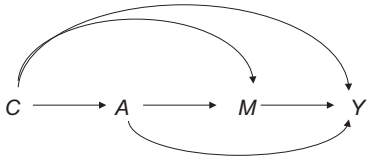

FIGURE 1. Causal diagram with exposure *A*, mediator *M*, outcome *Y*, and confounding variables *C*.

The estimation of direct and indirect effects in general requires stronger no-unmeasured-confounding assumptions than total effects. For causal diagrams interpreted as nonparametric structural equation models,[6] the following four assumptions suffice to identify natural direct and indirect effects from data[2]: (i) the effect of the exposure A on the outcome Y is unconfounded conditional on C; (ii) the effect of the mediator M on the outcome Y is unconfounded conditional on (C,A); (iii) the effect of the exposure A on the mediator M is unconfounded conditional on C; and (iv) none of the mediator-outcome confounders are affected by exposure. If we let $X \perp\!\!\!\perp Y|Z$ denote that X is independent of Y conditional on Z, then these four assumptions stated formally are (i) $Y_{am} \perp\!\!\!\perp A|C$, (ii) $Y_{am} \perp\!\!\!\perp M|\{A,C\}$, (iii) $M_a \perp\!\!\!\perp A|C$, and (iv) $Y_{am} \perp\!\!\!\perp M_{a^*}|C$.

Assumption (iv) is known as a cross-world independence assumption because it places an independence restriction on the joint distribution of the variables Y_{am} and M_{a^*}. The assumption states that knowing what would happen to the mediator when a person is unexposed, M_{a^*}, does not, within strata of covariates C, give information about the effects on the outcome of setting both the exposure and the mediator to certain other values, a and m, say (ie, does not give information on Y_{am}). It is a strong assumption because these two variables, Y_{am} and M_{a^*}, unlike the variables Y_{am} and M_a, are never observed together, and therefore, it is not possible to obtain empirical data (either experimental or observational) concerning their joint distribution. We could not confirm or disprove this assumption even if we could randomize both the exposure and the mediator. The cross-world independence assumption would hold on the causal diagram in Figure 1 interpreted as a nonparametric structural equation model of Pearl.[6] As noted above, however, there are other interpretations of causal diagrams than Pearl's.[6] The interpretation of causal diagrams of Robins,[7–9] in contrast to the nonparametric structural equation model of Pearl,[6] does not impose cross-world independencies; as a consequence, natural direct effects are not identified under the model (see Robins and Richardson[8] for further discussion). Nonetheless as mentioned several times, we will in general assume that causal diagrams are interpreted as a nonparametric structural equation models with independent errors throughout, though, as will be seen below, the final of our three approaches will not require this assumption.

Assumptions (i) and (ii) alone suffice to identify controlled direct effects. If these two assumptions hold, then the controlled direct effect is equal to

$$E\left[Y_{am} - Y_{a^*m}\right] = \Sigma_c \{E[Y|a, m, c] - E[Y|a^*, m, c]\}P(c).$$

If assumptions (i)–(iv) hold, then natural direct and indirect effects are identified and given by the following empirical expressions:

$$E\left[Y_{aM_{a^*}} - Y_{a^*M_{a^*}}\right] = \Sigma_{c,m}\{E[Y|a, m, c] - E[Y|a^*, m, c]\} P(m|a^*, c)P(c) \quad (1)$$

$$E\left[Y_{aM_a} - Y_{aM_{a^*}}\right] = \Sigma_{c,m} E[Y|a, m, c]\{P(m|a,c) - P(m|a^*,c)\}P(c). \quad (2)$$

VanderWeele and Vansteelandt[4,10] describe a simple regression-based method to estimate conditional natural direct effects, $E\left[Y_{aM_{a^*}} - Y_{a^*M_{a^*}}|c\right]$, and indirect effects, $E\left[Y_{aM_a} - Y_{aM_{a^*}}|c\right]$, and their standard errors from data (see also Imai et al[11] and Valeri and VanderWeele[12] for other estimation approaches and software). These expressions generalize those found in the social science literature[13,14] to allow for exposure-mediator interactions.

Even if a causal diagram is interpreted as a nonparametric structural equation model, assumption (iv) still requires that none of the mediator-outcome confounders are themselves affected by the exposure. This would hold in Figure 1 but would be violated in Figure 2. Avin et al[3] have shown that natural direct and indirect effects are not identified from data in Figure 2 or whenever there is a variable that is affected by exposure that in turn confounds the mediator-outcome relationship. To see why controlling just for L in a regression model will not suffice, suppose then that we are interested in estimating the direct effect of A on Y through pathways that do not involve M, and the mediated effect through pathways that do involve M. For the direct effect, we thus want to estimate the effects of two pathways $A - Y$ and $A - L - Y$ because these are the pathways not through M. Regression models will not work here because L is both a mediator-outcome confounder and on the pathway from the exposure to the outcome.

To intuitively see why regression control for L will not work, suppose first that we included L as a covariate in our outcome regression. In this case, we would potentially be blocking one of the direct-effect pathways not through M that we were interested in, namely $A - L - Y$, by controlling for L.

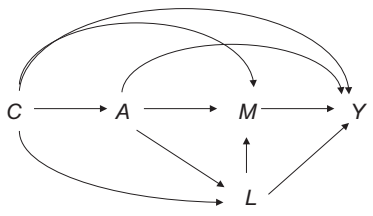

FIGURE 2. Causal diagram with a mediator-outcome confounding variable L that is affected by exposure.

This would suggest that we should perhaps not adjust for L in the regression analysis. However, if we do not adjust for L in the regression, then our estimates of the effect of M on Y will be confounded (because L is a confounder of the $M - Y$ relationship, which we have not controlled for), and thus, our direct and indirect effect estimates will be biased. Whether we control for L or not in the regression model, we will get bias if what we are interested in is the direct effect of A on Y not through M (but potentially through L). We get biased results if we adjust for L; we get biased results if we do not adjust for L. Simple regression methods cannot be used to estimate direct and indirect effects in this setting.

Controlled direct effects can, however, still be identified in the setting of Figure 2 but require methods other than regression in the presence of an exposure-induced mediator-outcome confounder. These methods have been described elsewhere.[15–19] In the next section, we describe three approaches to effect decomposition that are applicable even in the presence of an exposure-induced mediator-outcome confounder as in Figure 2. These approaches do not give the aforementioned natural direct and indirect effects, but they give other types of direct and indirect effects that may shed light on questions of mediation and pathways.

EFFECT DECOMPOSITION IN THE PRESENCE OF AN EXPOSURE-INDUCED MEDIATOR-OUTCOME CONFOUNDER

Approach 1: Joint Mediators

Consider the causal relationships in Figure 2. Suppose now that, instead of considering M only as the mediator, one were to consider (L, M) jointly as the mediator. We will need some additional notation. We let L_a denote the value of L that would have been observed had the exposure A been set to level a; we let Y_{alm} denote the value of the outcome that would have been observed had A been set to level a, L to l, and M to m. We could define the natural direct and indirect effects with (L, M) taken as the mediator as $E\left[Y_{aL_{a^*}M_{a^*}} - Y_{a^*L_{a^*}M_{a^*}}\right]$ and $E\left[Y_{aL_aM_a} - Y_{aL_{a^*}M_{a^*}}\right]$. The natural indirect effect here is the effect mediated through M or L or both, and the natural direct effect is the effect not through either M or L. We have the effect decomposition: $E\left[Y_a - Y_{a^*}\right] = E\left[Y_{aL_aM_a} - Y_{aL_{a^*}M_{a^*}}\right] + E\left[Y_{aL_{a^*}M_{a^*}} - Y_{a^*L_{a^*}M_{a^*}}\right]$.

The analog for four assumptions for identification can then be stated as: $(\text{i}^\dagger)$ $Y_{alm} \perp\!\!\!\perp A|C$, $(\text{ii}^\dagger)$ $Y_{alm} \perp\!\!\!\perp (L,M)|\{A,C\}$, $(\text{iii}^\dagger)$ $(L_a, M_a) \perp\!\!\!\perp A|C$, and $(\text{iv}^\dagger)$ $Y_{alm} \perp\!\!\!\perp (L_{a^*}, M_{a^*})|C$. In this case with ($L$, M) considered jointly as the mediator, assumption (iv) is once again effectively satisfied because on the causal diagram there is no effect of exposure A that confounds the relationship between the joint mediator (L, M) and the outcome Y. For assumptions (ii) and (iii), that there are no unmeasured "mediator-outcome" or "exposure-mediator" confounders,

respectively, these assumptions now need to apply to (*L, M*), considered jointly. Stated in intuitive terms, we thus need to control for confounders for the relationship between the exposure *A* and mediator *M*, along with confounders for the relationship between the exposure *A* and the variable *L*. Likewise, we need to control for confounders for the relationship between the mediator *M* and the outcome *Y*, along with confounders for the relationship between the variable *L* and the outcome *Y*.

Under the four assumptions above, the exact argument given in Pearl[2] for the identification of natural direct and indirect effects applies. Thus, these effects above are identified by

$$E\left[Y_{aL_{a^*}M_{a^*}} - Y_{a^*L_{a^*}M_{a^*}}\right] = \Sigma_{c,l,m}\{E[Y|a,l,m,c] - E[Y|a^*,l,m,c]\}P(l,m|a^*,c)P(c)$$

$$E\left[Y_{aL_aM_a} - Y_{aL_{a^*}M_{a^*}}\right] = \Sigma_{c,l,m}E[Y|a,l,m,c]\{P(l,m|a,c) - P(l,m|a^*,c)\}P(c).$$

Conditional analogs of these effects, $E\left[Y_{aL_{a^*}M_{a^*}} - Y_{a^*L_{a^*}M_{a^*}}|c\right]$ and $E\left[Y_{aL_aM_a} - Y_{aL_{a^*}M_{a^*}}|c\right]$, are given by the same expressions without summing over the distribution $P(c)$. In the following section, we will describe a weighting-based estimator for these effects.

Approach 2: Path-specific Effects

If *M* is the actual mediator of interest, the approach above may be unsatisfactory because it gives only mediated effects when *M* and *L* are considered jointly as the mediator. Let us return to the setting where the interest is principally in *M* as a mediator, but such that there is an exposure-induced mediator-outcome confounder *L* as in Figure 2. Although natural direct and indirect effects with *M* considered alone as the mediator are not in general identified under the causal diagram of Figure 2, certain path-specific effects are identified. For example, although we cannot identify the effects mediated through pathways involving *M* (ie, the combination of $A \rightarrow L \rightarrow M \rightarrow Y$ and $A \rightarrow M \rightarrow Y$) and the effects through pathways not involving *M* (ie, the combination of $A \rightarrow Y$ and $A \rightarrow L \rightarrow Y$), Avin et al[3] showed that we can identify the effects (1) through pathways involving neither *L* nor *M* (ie, $A \rightarrow Y$), (2) through the additional pathways not involving *L* (i.e. $A \rightarrow M \rightarrow Y$), and (3) through the pathways involving only *L* (ie, the combination of $A \rightarrow L \rightarrow M \rightarrow Y$ and $A \rightarrow L \rightarrow Y$). For simplicity, let us refer to these effects as $E_{A\rightarrow Y}$, $E_{A\rightarrow M\rightarrow Y}$, and $E_{A\rightarrow LY}$. In counterfactual notation, these effects are

$$E_{A\rightarrow Y}(c) = E\left[Y_{aL_{a^*}M_{a^*}} - Y_{a^*L_{a^*}M_{a^*}}|c\right],$$

$$E_{A\rightarrow M\rightarrow Y}(c) = E\left[Y_{aL_{a^*}M_{aL_{a^*}}} - Y_{aL_{a^*}M_{a^*}}|c\right],$$

and

$$E_{A\rightarrow LY}(c) = E\left[Y_{aL_aM_a} - Y_{aL_{a^*}M_{aL_{a^*}}}|c\right]$$

(see the eAppendix, http://links.lww.com/EDE/A744, for further discussion). In addition, as shown in the eAppendix (http://links.lww.com/EDE/A744), a total effect decomposes into the sum of these three effects: $E[Y_a - Y_{a^*}] = E_{A\rightarrow Y} + E_{A\rightarrow M\rightarrow Y} + E_{A\rightarrow LY}$. By the results of Avin et al,[3] it follows that if Figure 2 is a causal diagram (and thus the identifying assumptions (i†)-(iv†) all hold), then these three effects are identified and in fact given by the following empirical expressions:

$$E_{A\rightarrow Y} = \Sigma_{c,l,m}\{E[Y|c,a,l,m] - E[Y|c,a^*,l,m]\} P(l,m|a^*,c)P(c)$$

$$E_{A\rightarrow M\rightarrow Y} = \Sigma_{c,l,m}E[Y|c,a,l,m] \{P(m|c,a,l) - P(m|c,a^*,l)\}P(l|c,a^*)P(c)$$

$$E_{A\rightarrow LY} = \Sigma_{c,l,m}E[Y|c,a,l,m] P(m|c,a,l)\{P(l|c,a) - P(l|c,a^*)\}P(c).$$

Conditional analogs of these effects, $E_{A\rightarrow Y}(c)$, $E_{A\rightarrow M\rightarrow Y}(c)$, and $E_{A\rightarrow LY}(c)$, are given by the same expressions without summing over the distribution $P(c)$. Note that these expressions do not allow us to distinguish between effects through *L* and through *M* versus those that are through *L* but not through *M* (see the eAppendix, http://links.lww.com/EDE/A744, for further discussion and formality). Once again, in the following section, we will describe weighting-based estimators for these effects.

Approach 3: Interventional Effects

Suppose again that we wish to retain *M* as our principal mediator, rather than *M* and *L* jointly. Although natural direct and indirect effects with *M* alone as the mediator of interest are not identified, alternative effects that randomly set *M* to a value chosen from the distribution of a particular exposure level can be identified. Let $G_{a|C}$ denote a random draw from the distribution of the mediator among those with exposure status *a* conditional on *C*.

Suppose *a* and *a** are two values of the exposure we wish to compare. The effect $E(Y_{aG_{a|C}}) - E(Y_{aG_{a^*|C}})$ is then the effect on the outcome of randomly assigning a person who is given the exposure to a value of the mediator from the distribution of the mediator among those given exposure versus no exposure (given covariates); this is an effect through the mediator. Next, consider the effect $E(Y_{aG_{a^*|C}}) - E(Y_{a^*G_{a^*|C}})$; this is a direct effect comparing exposure versus no exposure with the mediator in both cases randomly drawn from the distribution of the population when given no exposure (given covariates). Finally, the effect $E(Y_{aG_{a|C}}) - E(Y_{a^*G_{a^*|C}})$ compares the expected outcome when having the exposure, with the mediator randomly drawn from the distribution of the population

when given the exposure (given covariates) to the expected outcome when not having the exposure, with the mediator randomly drawn from the distribution of the population when not exposed. These various effects are similar to those described by Didelez et al[20] and Geneletti.[21] With effects thus defined we have the decomposition:

$$E(Y_{aG_{a|C}}) - E(Y_{a^*G_{a^*|C}}) = \{E(Y_{aG_{a|C}}) - E(Y_{aG_{a^*|C}})\} + \{E(Y_{aG_{a^*|C}}) - E(Y_{a^*G_{a^*|C}})\},$$

so that the overall effect decomposes into the sum of the effect through the mediator and the direct effect. These are not the natural direct and indirect effects considered earlier but are instead analogs arising from not fixing the mediator for each person to the level it would have been under a particular exposure, but rather to a level that is randomly chosen from the distribution of the mediator among all those with a particular exposure.

To identify these effects, the following conditions suffice: assumptions (i) $Y_{am} \perp\!\!\!\perp A \mid C$ and (iii) $M_a \perp\!\!\!\perp A \mid C$ above, that conditional on C there is no unmeasured exposure-outcome or exposure-mediator confounding, along with an assumption (ii*) that $Y_{am} \perp\!\!\!\perp M \mid \{A, C, L\}$, that is, that conditional on (A, C, L), there is no unmeasured confounding of the mediator-outcome relationship. These three assumptions would hold in the causal diagram in Figure 2, even if the association between L and Y was confounded by unmeasured factors. If these three assumptions hold, then these interventional effects are identified by

$$E(Y_{aG_{a^*|C}}) - E(Y_{a^*G_{a^*|C}}) = \Sigma_{c,l,m}\{E[Y \mid a,l,m,c]P(l \mid a,c) - E[Y \mid a^*,l,m,c]P(l \mid a^*,c)\} \times P(m \mid a^*,c)P(c)$$

$$E(Y_{aG_{a|C}}) - E(Y_{aG_{a^*|C}}) = \Sigma_{c,l,\boldsymbol{m}} E[Y \mid a,l,m,c]P(l \mid a,c) \{P(m \mid a,c) - P(m \mid a^*,c)\}P(c),$$

where the first expression amounts to averaging controlled direct effects corresponding to different levels m, using the distribution $P(m \mid a^*, c)$. Conditional analogs of these effects are given by the same expressions without summing over the distribution $P(c)$. These expressions reduce to the mediation formulae (1) and (2) when L does not confound the association between M and Y, conditional on covariates C. Note that in contrast to the effects described in the first two approaches, the effects here in the third approach do not require interpreting causal diagrams as nonparametric structural equation models; the effects are identified under alternative interpretations[7–9] because assumptions (i), (ii), and (iii), none of which are "cross-world," are implied by the models. Once again, in the following section, we will describe weighting-based estimators for these effects.

WEIGHTING ESTIMATORS

All weighting-based estimators require first estimating inverse-probability weights, which can be obtained on the basis of regression models for the exposure A, mediator M, and confounder L. For pedagogic purposes, we will focus on dichotomous (coded 0 or 1) exposure, mediator and confounder, in which case three logistic regressions can be used: (1) a logistic regression of the exposure A on covariates, (2) a logistic regression of the confounder L on the exposure and covariates, and (3) a logistic regression of the mediator on the confounder, exposure, and covariates. We also focus on weighting estimators for the marginal effects described above. Analogous estimators for the conditional effects are described in the eAppendix (http://links.lww.com/EDE/A744). Under approach 1, a weighting-based estimator can then be obtained upon duplicating the data set and adding an exposure variable A^* that is 0 for the first replication and 1 for the second. For each person, a weight is obtained by taking the product of the predicted probabilities (of the observed confounder and mediator values) from the two logistic regressions for L and M had the exposure been A^*, divided by the product of the corresponding predicted probabilities from the two logistic regressions had the exposure been as observed, with additional weighting by the reciprocal of the probability of the observed exposure to adjust for confounding of the exposure-outcome association:

$$\frac{P(l \mid a^*,c)P(m \mid l,a^*,c)}{P(a \mid c)P(l \mid a,c)P(m \mid l,a,c)}.$$

The natural direct effect $E[Y_{1L_0M_0} - Y_{0L_0M_0}]$ is then obtained as the coefficient of A in a weighted regression of Y on A among persons with $A^* = 0$; the natural indirect effect $E[Y_{1L_1M_1} - Y_{1L_0M_0}]$ is obtained as the coefficient of A^* in a weighted regression of Y on A^* among persons with $A = 1$ in the duplicated data set. Approach 3 works like the first, but using the weights

$$\frac{\Sigma_l P(m \mid l,a^*,c)P(l \mid a^*,c)}{P(a \mid c)P(m \mid l,a,c)}$$

instead.

Under approach 2, a weighting-based estimator can be obtained upon merging three copies of the data set and adding exposure variables A^* and A^{**}, where A^* equals the observed exposure for the first replication and $1 - A$ for the next two replications, and where A^{**} equals the observed exposure for the first two replications and $1 - A$ for the third replication. For each person, a weight is now obtained by taking the product of the predicted probability (of the observed confounder value) from the logistic regression for L had the exposure been A^* and the predicted probability (of the observed mediator value) from the logistic regression for M had the exposure been A^{**}, divided by the product of the corresponding predicted probabilities from the two logistic regressions had the exposure

been as observed, with additional weighting by the reciprocal of the probability of the observed exposure:

$$\frac{P(l \mid a^*, c)P(m \mid l, a^{**}, c)}{P(a \mid c)P(l \mid a, c)P(m \mid l, a, c)}.$$

For a binary exposure, the natural direct effect $E_{A\rightarrow Y}$ is now obtained as the coefficient of A in a weighted regression of Y on A among persons with $A^* = A^{**} = 0$; the natural indirect effect $E_{A\rightarrow LY}$ is obtained as the coefficient of A^* in a weighted regression of Y on A^* among persons with $A = 1, A^{**} = 0$; finally, the natural indirect effect $E_{A\rightarrow M\rightarrow Y}$ is obtained as the coefficient of A^{**} in a weighted regression of Y on A^{**} among those with $A = A^* = 1$.

SAS code (SAS Institute, Cary, NC) for each of these three weighting approaches are given in the eAppendix (http://links.lww.com/EDE/A744). Because the procedures above are not maximum likelihood procedures, involve creating copies of the observed data, and moreover require estimation of the weights, the standard confidence intervals and estimates of the standard error can be severely biased. Valid standard errors and confidence intervals can be obtained via the bootstrap.

ILLUSTRATION

To illustrate the three approaches, we will analyze 2003 US birth certificate data and will consider whether the effect of the exposure, *A*, of adequate or inadequate prenatal care (n = 2,629,247; excluding those with intermediate or superadequate care for the purposes of this illustration) on preterm birth (*Y*) is mediated by preeclampsia (*M*) or other pathways, where maternal smoking (*L*) is taken as an exposure-induced mediator-outcome confounder. Maternal smoking may be affected by prenatal care and may in turn affect both preeclampsia and preterm birth. The analysis is principally for the purpose of illustration. Categories of the adequacy of prenatal care are determined from data on the month prenatal care was initiated, on the number of visits, and on gestational age, according to the American College of Gynecologists recommendation, as encoded in a modification of the Adequacy of Prenatal Care Utilization index.[22,23] In this analysis, we will take age category (<20 years, between 20 and 35 years, or >35 years), ethnicity (black, Hispanic, Native American, white), education, and marital status as baseline confounders (*C*). Our analysis is certainly a simplification of a more complex reality, in that prenatal care and maternal smoking are both ultimately time-varying, and preeclampsia and preterm birth could be regarded as processes, whereas we will treat them as dichotomous; again, the analysis here is used only for illustrative purposes.

Inverse probability weights were constructed on the basis of logistic regression models for adequate care, smoking, and preeclampsia. In view of the large sample size and the resulting computational complexity, standard errors and confidence intervals were constructed using the subsampling bootstrap methods.[24] This is similar to the bootstrap approach but involved repeating the analysis for 1000 subsamples of size n = 13,146 (0.5% of the total sample size); on the basis of the empirical standard deviation of the 1000 estimates, the standard error of the estimates that were obtained from the analysis of the full data set can be inferred (accounting for correlation resulting from the fact that some data points may be shared between subsamples).

Approach 1, which considers both preeclampsia and smoking as mediators, shows that the conditional direct effect of adequate care, other than via smoking or preeclampsia, is to reduce the odds of preterm birth by 54.3% (95% confidence interval = 53.9% to 54.8%). The remaining indirect effect via smoking or preeclampsia amounts to a 0.7% (0.6% to 0.8%) reduction i the odds of preterm birth. Approach 2 gives the same conditional direct effect, $E_{A\rightarrow Y}$, but is more informative about the mediated effect. It indicates that the indirect effect via smoking, $E_{A\rightarrow LY}$, amounts to a 0.7% (0.7% to 0.8%) reduction in the odds of preterm birth. That this effect is small is perhaps not surprising because the effect of adequate care by decreasing smoking mixes an inherent beneficial impact on preterm birth with a harmful effect by increasing preeclampsia (because smoking prevents preeclampsia). The remaining indirect effect via preeclampsia, but not smoking, $E_{A\rightarrow M\rightarrow Y}$, is to increase the odds of preterm birth by 0.06% (−0.01% to 0.12%). In contrast to the previous two approaches, approach 3 avoids assumptions of unconfoundedness with respect to smoking. It suggest that the conditional direct effect of adequate care, other than via preeclampsia, is to reduce the odds of preterm birth by 54.7% (54.3% to 55.2%), corresponding to a negligible remaining indirect effect via preeclampsia that amounts to a reduction in the odds of preterm birth with 0.8% (0.7% to 0.9%). The inverse probability weights for all analyses were very stable, varying between 0.70 and 1.43 in approaches 1 and 2, and 0.80 and 1.25 in approach 3.

DISCUSSION

We have described three approaches to effect decomposition in the presence of an exposure-induced mediator-outcome confounder. This setting presents challenges to the counterfactual approach to mediation because natural direct and indirect effects are not identified. We considered approaches that estimate (1) natural direct and indirect effects with the exposure-induced confounder and the original mediator of interest considered jointly as the mediator, (2) certain path-specific effects that are identified from the data, or (3) analogs to the notions of natural direct and indirect effects that instead rely on randomized interventions on the mediator. The three approaches estimate different quantities and may be of interest in different contexts. All three can, however, potentially be pursued in any particular application and, considered together, may shed considerable light on questions of mediation and pathways.

It should also be noted that even with the three approaches described in this article, the identification of these

effects requires strong no-unmeasured-confounding assumptions. Sensitivity analysis techniques have been developed for natural direct and indirect effects when there is no exposure-induced mediator-outcome confounding.[11,25] Similar sensitivity analysis techniques could potentially be developed for each of the three approaches considered in this article. However, until such approaches have been developed, an analyst must proceed cautiously when evaluating these assumptions, when attempting to control for sufficient covariates to make these plausible, and when interpreting the results as providing only tentative evidence. Further development of sensitivity analysis techniques for the approaches in this article will allow for more formal evaluation of the extent to which substantive conclusions may hold even when assumptions are violated.

An alternative approach to reasoning about mediation in the context of an exposure-induced mediator-outcome confounder would be instead to maintain the original natural direct and indirect effects as the targets of interest and use sensitivity analysis to help address issues of nonidentifiability. Natural direct and indirect effects are not identified in this setting, but sensitivity analysis approaches can be useful in assessing the extent to which various estimators using the observed data can deviate from the true natural direct and indirect effects. Several approaches along these lines have begun to develop. For example, Imai and Yamamoto[26] proposed a parametric sensitivity analysis technique for linear models to reason about natural direct and indirect in the presence of an exposure-induced mediator-outcome confounder, which requires data on the exposure-induced mediator-outcome confounder. Their technique is at present, however, restricted to one fairly simple setting. Tchetgen Tchetgen and Shpitser,[27] and VanderWeele and Chiba[28] proposed nonparametric techniques that are more general and do not require data on the exposure-induced mediator-outcome confounder but require specifying a larger number of sensitivity analysis parameters that may be difficult to do in practice. Vansteelandt and VanderWeele[29] describe a technique that, like that of Imai and Yamamoto,[26] requires data to be available on the exposure-induced mediator-outcome confounder and also requires specifying a selection bias function, which can be difficult to interpret in practice, but does have the advantage that the selection bias function is essentially zero as long as there is no three-way interaction among the exposure, mediator, and exposure-induced confounder. These various sensitivity techniques could be used in conjunction with the methods described there to improve conclusions about mediation and pathways.

REFERENCES

1. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology*. 1992;3:143–155.
2. Pearl J. Direct and indirect effects. In: Breese O, Koller D, eds. *Proceedings of the Seventeenth Conference on Uncertainty and Artificial Intelligence*. San Francisco, CA: Morgan Kaufmann; 2001:411–420.
3. Avin C, Shpitser I, Pearl J. Identifiability of path-specific effects. In: *Proceedings of the International Joint Conferences on Artificial Intelligence.* 2005;357–363.
4. VanderWeele TJ, Vansteelandt S. Conceptual issues concerning mediation, interventions and composition. *Stat Interface*. 2009;2:457–468.
5. VanderWeele TJ. Concerning the consistency assumption in causal inference. *Epidemiology*. 2009;20:880–883.
6. Pearl J. *Causality: Models, Reasoning, and Inference*. 2nd ed. Cambridge: Cambridge University Press; 2009.
7. Robins JM. Semantics of causal DAG models and the identification of direct and indirect effects. In: Green P, Hjort NL, Richardson S, eds. *Highly Structured Stochastic Systems*. New York, NY: Oxford University Press; 2003:70–81.
8. Robins JM, Richardson TS. Alternative graphical causal models and the identification of direct effects. In: Shrout P, ed. *Causality and Psychopathology: Finding the Determinants of Disorders and Their Cures*. New York, NY: Oxford University Press; 2010:427–437.
9. Robins JM. A new approach to causal inference in mortality studies with sustained exposure period—application to control of the healthy worker survivor effect. *Math Model*. 1986;7:1393–1512.
10. Vanderweele TJ, Vansteelandt S. Odds ratios for mediation analysis for a dichotomous outcome. *Am J Epidemiol*. 2010;172:1339–1348.
11. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. *Psychol Methods*. 2010;15:309–334.
12. Valeri L, Vanderweele TJ. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. *Psychol Methods*. 2013;18:137–150.
13. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol*. 1986;51:1173–1182.
14. MacKinnon DP. *An Introduction to Statistical Mediation Analysis*. New York: Lawrence Erlbaum Associates; 2008.
15. van der Laan MJ, Petersen ML. Direct effect models. *Int J Biostat*. 2008;4:Article 23.
16. Robins JM. Testing and estimation of direct effects by reparameterizing directed acyclic graphs with structural nested models. In: Glymour C, Cooper GF, eds. *Computation, Causation, and Discovery*. Menlo Park, CA, Cambridge, MA: AAAI Press/The MIT Press; 1999:349–405.
17. Goetgeluk S, Vansteelandt S, Goetghebeur E. Estimation of controlled direct effects. *J Royal Stat Soc*. 2008;70:1049–1066.
18. VanderWeele TJ. Marginal structural models for the estimation of direct and indirect effects. *Epidemiology*. 2009;20:18–26.
19. Vansteelandt S. Estimating direct effects in cohort and case-control studies. *Epidemiology*. 2009;20:851–860.
20. Didelez V, Dawid AP, Geneletti S. Direct and indirect effects of sequential treatments. In: Dechter R, Richardson TS, eds. *Proceedings of the 22nd Annual Conference on Uncertainty in Artificial Intelligence*. Arlington, VA: AUAI Press; 2006:138–146.
21. Geneletti S. Identifying direct and indirect effects in a non-counterfactual framework. *J Royal Stat Soc*. 2007;69:199–216.
22. Kotelchuck M. An evaluation of the Kessner Adequacy of Prenatal Care Index and a proposed Adequacy of Prenatal Care Utilization Index. *Am J Public Health*. 1994;84:1414–1420.
23. VanderWeele TJ, Lantos JD, Siddique J, Lauderdale DS. A comparison of four prenatal care indices in birth outcome models: comparable results for predicting small-for-gestational-age outcome but different results for preterm birth or infant mortality. *J Clin Epidemiol*. 2009;62:438–445.
24. Politis, DN, Romano JP. Large sample confidence regions based on subsamples under minimal assumptions. *Ann Stat*. 1994;22:2031–2050.
25. VanderWeele TJ. Bias formulas for sensitivity analysis for direct and indirect effects. *Epidemiology*. 2010;21:540–551.
26. Imai K, Yamamoto T. Identification and sensitivity analysis for multiple causal mechanisms: revisiting evidence from framing experiments. *Political Anal*. 2012;21:141–171.
27. Tchetgen Tchetgen EJ, Shpitser I. Semiparametric theory for causal mediation analysis: efficiency bounds, multiple robustness and sensitivity analysis. *Ann Stat*. 2012;40:1816–1845.
28. VanderWeele TJ, Chiba, Y. Sensitivity analysis for direct and indirect effects in the presence of an mediator-outcome confounders that may be affected by exposure. *Epidemiol Biostat Public Health*. In press.
29. Vansteelandt S, Vanderweele TJ. Natural direct and indirect effects on the exposed: effect decomposition under weaker assumptions. *Biometrics*. 2012;68:1019–1027.